Stationary Probability Distributions of a Markov Chain

Berlian Setiawaty
Department of Mathematics, Bogor Agricultural University
Jl. Raya Pajajaran Bogor, 16144 Indonesia

Abstract

This article shows that stationary probability distributions of a Markov chain can be classified into two classes. These classes are determined by the type of communicating classes of the chain.

Keywords : Markov chain, Stationary probability distributions, Communicating classes

1 Introduction

The aim of this article is to show the relations between communicating classes and stationary probability distributions of a Markov chain. By analysing the type of the communicating classes of the chain, we will show that stationary probability distributions can be classified into two classes.

In section 2, definitions and basic result regarding Markov chain are given. The main result of the article is presented in section 3.

2 Markov Chains

Let $\{X_t : t \in \mathbb{N}\}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$, taking values in a finite set $S = \{1, \ldots, K\}$. $\{X_t\}$ is said to be a *Markov chain* if it satisfies

$$P(X_{n+1} = i_{n+1} | X_1 = i_1, \ldots, X_n = i_n) = P(X_{n+1} = i_{n+1} | X_n = i_n), \qquad (1)$$

for all $i_1, \ldots, i_{n+1} \in S$ and $n \in \mathbb{N}$. Property (1) is called the *Markov property*.

Let $m \leq n$, then by (1),

$$
\begin{aligned}
&P(X_{n+1} = i_{n+1} | X_1 = i_1, \dots, X_m = i_m) \\
&= \sum_{i_{m+1}=1}^{K} \cdots \sum_{i_n=1}^{K} P(X_{m+1} = i_{m+1}, \dots, X_{n+1} = i_{n+1} | X_1 = i_1, \dots, X_m = i_m) \\
&= \sum_{i_{m+1}=1}^{K} \cdots \sum_{i_n=1}^{K} \Big\{ P(X_{m+1} = i_{m+1} | X_1 = i_1, \dots, X_m = i_m) \\
&\qquad \times P(X_{m+2} = i_{m+2} | X_1 = i_1, \dots, X_{m+1} = i_{m+1}) \\
&\qquad \times \cdots \times P(X_{n+1} = i_{n+1} | X_1 = i_1, \dots, X_n = i_n) \Big\} \\
&= \sum_{i_{m+1}=1}^{K} \cdots \sum_{i_n=1}^{K} \Big\{ P(X_{m+1} = i_{m+1} | X_m = i_m) \cdot P(X_{m+2} = i_{m+2} | X_{m+1} = i_{m+1}) \\
&\qquad \times \cdots \times P(X_{n+1} = i_{n+1} | X_n = i_n) \Big\} \\
&= \sum_{i_{m+1}=1}^{K} \cdots \sum_{i_n=1}^{K} P(X_{m+1} = i_{m+1}, \dots, X_{n+1} = i_{n+1} | X_m = i_m) \\
&= P(X_{n+1} = i_{n+1} | X_m = i_m). \qquad (2)
\end{aligned}
$$

So the Markov property (1) is equivalent with (2).

Assume that $P(X_{n+1} = j | X_n = i)$ depends only on (i, j) and not on n. Let

$$\alpha_{ij} = P(X_{n+1} = j | X_n = i), \qquad i, j = 1, 2, \dots, K, \tag{3}$$

then α_{ij} are called the *transition probabilities* from state i to state j and the $K \times K$ matrix A defined by

$$A = (\alpha_{ij}), \tag{4}$$

is called the *transition probability matrix* of the Markov chain $\{X_t\}$. Notice that A satisfies

$$
\begin{array}{ll}
0 \leq \alpha_{ij} \leq 1, & i, j = 1, \dots, K \\
\sum_{j=1}^{K} \alpha_{ij} = 1, & i = 1, \dots, K.
\end{array}
$$

Thus A is a *stochastic matrix.*

Let

$$\pi_i = P(X_1 = i), \qquad i = 1, \dots, K \tag{5}$$

and

$$\pi = (\pi_i). \tag{6}$$

The $1 \times K$-matrix π is called the *initial probability distribution* of the Markov chain $\{X_t\}$. Notice that π satisfies

$$0 \le \pi_i \le 1, \qquad i = 1, \dots, K \qquad \text{and} \qquad \sum_{i=1}^{K} \pi_i = 1.$$

By (1), (3), (4), (5) and (6),

$$\begin{aligned} P(X_1 = i_1, \dots, X_n = i_n) &= P(X_1 = i_1) \cdot P(X_2 = i_2 | X_1 = i_1) \\ &\quad \times \dots \times P(X_n = i_n | X_{n-1} = i_{n-1}) \\ &= \pi_{i_1} \cdot \alpha_{i_1, i_2} \cdots \alpha_{i_{n-1}, i_n}. \end{aligned} \tag{7}$$

Then by (7),

$$\begin{aligned} P(X_n = i) &= \sum_{i_1=1}^{K} \cdots \sum_{i_{n-1}=1}^{K} P(X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i_n) \\ &= \sum_{i_1=1}^{K} \cdots \sum_{i_{n-1}=1}^{K} \pi_{i_1} \cdot \alpha_{i_1, i_2} \cdots \alpha_{i_{n-1}, i} \\ &= \pi A^{n-1} e_i, \end{aligned} \tag{8}$$

where $A^n = AA \cdots A$ and $e_i = (0, \dots, 0, 1, 0, \dots, 0)^T$. Hence, it can be concluded that the probability distribution of the Markov chain $\{X_t\}$ is completely determined by the initial probability π and the transition probability matrix A.

If the initial probability distribution π satisfies

$$\pi A = \pi, \tag{9}$$

then π is called a *stationary probability distribution* with respect to A. By (8) and (9), for every $n \in \mathbf{N}$,

$$\begin{aligned} P(X_n = i) &= \pi A^{n-1} e_i \\ &= \pi e_i \\ &= \pi_i, \end{aligned}$$

implying

$$\begin{aligned} P(X_{m+1} = i_1, \dots, X_{m+n} = i_n) &= P(X_{m+1} = i_1) \cdot P(X_{m+2} = i_2 | X_{m+1} = i_1) \\ &\quad \times \dots \times P(X_{m+n} = i_n | X_{m+n-1} = i_{n-1}) \\ &= \pi_{i_1} \cdot \alpha_{i_1, i_2} \cdots \alpha_{i_{n-1}, i_n} \\ &= P(X_1 = i_1, \dots, X_n = i_n), \end{aligned}$$

for $m \in \mathbf{N}$ and $i_1, \dots, i_n \in S$. So in this case, the Markov chain $\{X_t\}$ is *(strictly) stationary.*

Stationary Probability Distributions of a Markov Chain

To classify the states of the Markov chain $\{X_t\}$, define a *communication* relation "$\leftrightarrow$" as follows. A state j is said to be *accessible* or reachable from a state i, denoted as $i \to j$, if there is an integer n, $0 \le n < K$, such that the (i,j) entry of A^n is positive. If $i \to j$ and $j \to i$, then i and j are said to *communicate* with each other, denoted as $i \leftrightarrow j$.

For each state i, define a *communicating class*

$$C(i) = \{j \in S : i \leftrightarrow j\}.$$

Since relation $\leftrightarrow$ is an equivalence relation, then the communicating classes satisfy :

(a). For every state i, $i \in C(i)$.
(b). If $j \in C(i)$, then $i \in C(j)$.
(c). For any state i and j, either $C(i) = C(j)$ or $C(i) \cap C(j) = \emptyset$
Thus the state space S can be *partitioned* into these classes.

A Markov chain is said to be *irreducible*, if all states communicate with each other. So in this case, the Markov chain has only one communicating class.

A communicating class C is called *ergodic* if

$$\sum_{j \in C} \alpha_{ij} = 1, \qquad \forall i \in C. \tag{10}$$

The individual states in an ergodic class are also called *ergodic*.

A communicating class C is called *transient*, if there is $i \in C$, such that

$$\sum_{j \in C} \alpha_{ij} < 1. \tag{11}$$

The individual states in a transient class are also called *transient*.

To identify the transition probability matrix within a communicating class, the irreducibility of square matrix is introduced. An $n \times n$-matrix $B = (\beta_{ij})$ is said to be *irreducible*, if there is a permutation of indices σ, such that the matrix $\widetilde{B} = (\widetilde{\beta}_{ij})$, with $\widetilde{\beta}_{ij} = \beta_{\sigma(i),\sigma(j)}$, has form

$$\widetilde{B} = \begin{pmatrix} C & 0 \\ D & E \end{pmatrix}$$

where C and E are $l \times l$ and $m \times m$ matrices respectively, and $l + m = n$.

Let C_e be an ergodic class and n_e be the number of ergodic states in C_e. Let A_e be the $n_e \times n_e$ transition probability matrix within C_e. Then by (10) A_e is a *stochastic* matrix. Moreover, A_e is *irreducible*, since if A_e is *reducible*, then by some permutation σ, A_e can be reduced to the form

$$\widetilde{A}_e = \begin{pmatrix} B_e & 0 \\ C_e & D_e \end{pmatrix}, \tag{12}$$

where B_e and D_e are $k_e \times k_e$ and $l_e \times l_e$ matrices respectively, with $k_e + l_e = n_e$. But from (12), it can be seen that every state in $\{\sigma(1), \ldots, \sigma(k_e)\}$ does not communicate with every state in $\{\sigma(k_e+1), \ldots, \sigma(n_e)\}$, contradicting with the fact that C_e is a communicating class. Therefore, A_ϵ must not be reducible.

Let C_t be a transient class and n_t be the number of transient states in C_t. Let A_t be the $n_t \times n_t$ transition probability matrix within C_t. Then by (11), A_t is a *substochastic* matrix, that is, its individual row sums are ≤ 1.

3 Relations between Communicating Classes and Stationary Probability Distributions of a Markov Chain

The next lemma shows the relation between irreducible Markov chains and irreducible transition probability matrices.

Lemma 1 *Let $\{X_t\}$ be a Markov chain with a $K \times K$ transition probability matrix A. Then $\{X_t\}$ is irreducible if and only if A is irreducible.*

Proof :

Let $\{X_t\}$ be a Markov chain with a $K \times K$ transition probability matrix A. If $\{X_t\}$ is irreducible, then it consists of a single communicating class C and the transition probablity matrix within C is A. Since A is a stochastic matrix, then C is an ergodic class. From the ergodicity of C, the irreducibility of A follows.

On the otherhand, if A is irreducible, then from [1], page 63, for every $1 \leq i, j \leq K$, there is an integer n, $0 \leq n \leq K$, such that the (i,j) entry of A^n is positive. This means that every state communicates with each other. So the chain $\{X_t\}$ is irreducible. ∎

Let $\{X_t\}$ be a Markov chain with a $K \times K$ transition probability matrix A. Let K_e and K_t be the number of ergodic states and transient states respectively. In general, after a suitable permutation of indices, the transition probability matrix A can be written in the block form as

$$\tilde{A} = \begin{pmatrix} B & 0 \\ C & D \end{pmatrix},$$

where B is a $K_e \times K_e$-stochastic matrix and D is a $K_t \times K_t$-substochastic matrix.

The block D describes the transient $\rightarrow$ transient movements in the chain. For each class of transient states, at least one row in D will have sum < 1.

The $K_t \times K_e$-block C describes the transient $\rightarrow$ ergodic movements in the chain. For each class of transient states, at least one row in C will have a non-zero entry.

Finally, The $K_e \times K_e$-block B describes the movements within each ergodic class in the chain. Suppose that the chain has e ergodic classes. Since it is impossible to leave an ergodic class, B has the form,

$$B = \begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_e \end{pmatrix},$$

where B_i is the transition matrix within the i-th ergodic class. For each i, B_i is an irreducible stochastic matrix.

The following lemma shows the relation between the communicating classes and the stationary probability distributions.

Lemma 2 *Let $\{X_t\}$ be a Markov chain with a $K \times K$ transition probability matrix A. Let*

$$\begin{aligned} S &= \{\pi = (\pi_i) \,:\, \pi_i \geq 0,\ i = 1, \ldots, K,\ \sum_{i=1}^{K} \pi_i = 1,\ \pi A = \pi\} \\ S^+ &= \{\pi \in S \,:\, \pi_i > 0,\ i = 1, \ldots, K\} \\ S^o &= S - S^+. \end{aligned}$$

1. *If $\{X_t\}$ is irreducible, then $S = S^+$ and $S^o = \emptyset$.*
2. *If $\{X_t\}$ has e communicating classes, with $2 \leq e \leq K$ which all are ergodic, then $S^+ \neq \emptyset$ and $S^o \neq \emptyset$.*
3. *If $\{X_t\}$ has k communicating classes, with $2 \leq k \leq K$, where e of them are ergodic, $1 \leq e < k$, and t of them are transient, $e + t = k$, then $S = S^o$ and $S^+ = \emptyset$.*

Proof :
To prove (a), let $\{X_t\}$ be an irreducible Markov chain. Suppose there is $\pi \in S^o$. Let k be the number of non-zero π_i. Without loss of generality, suppose that

$$\begin{aligned} \pi_i &> 0, \quad \text{for} \quad i = 1, \ldots, k \\ \pi_i &= 0, \quad \text{for} \quad i = k+1, \ldots, K. \end{aligned}$$

As $\pi A = \pi$, then

$$\alpha_{ij} = 0, \qquad \text{for} \quad i = 1, \ldots, K, \qquad j = k+1, \ldots, K.$$

Thus A has form

$$A = \begin{pmatrix} B & 0 \\ C & D \end{pmatrix},$$

where B is a $k \times k$-matrix and D is a $(K-k) \times (K-k)$-matrix. So A is reducible, contradicting with the fact that A is irreducible by Lemma 1. Therefore, it must be $S^o = \emptyset$ and hence $S = S^+$.

To prove (b), let $\{X_t\}$ be a Markov chain having e communicating classes with $2 \le e \le K$, which all are ergodic. Then without loss of generality, A is of the form

$$A = \begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_e \end{pmatrix},$$

where B_i is the transition matrix within the i-th ergodic class and it is an irreducible stochastic matrix.

Let π^i be an $1 \times e_i$-matrix, where e_i is the number of ergodic states in B_i, such that for $i = 1, \ldots, e$,

$$\pi^i_j \ge 0, \qquad j = 1, \ldots, e_i, \qquad \text{and} \qquad \sum_{j=1}^{e_i} \pi^i_j = 1$$

and

$$\pi^i B_i = \pi^i.$$

By (a),

$$\pi^i_j > 0, \qquad \text{for} \quad j = 1, \ldots, e_i \qquad \text{and} \quad i = 1, \ldots, e.$$

Let

$$\widehat{\pi} = (\pi^1, 0, \ldots, 0),$$

then

$$\begin{aligned} \widehat{\pi} A &= (\pi^1 B_1, 0, \ldots, 0) \\ &= (\pi^1, 0, \ldots, 0) \\ &= \widehat{\pi}. \end{aligned}$$

So $\widehat{\pi} \in S^o$ and hence $S^o \neq \emptyset$.

Let a_i, $i = 1, \ldots, e$ be any real numbers such that

$$a_i > 0, \qquad i = 1, \ldots, e \qquad \text{and} \qquad \sum_{i=1}^{e} a_i = 1.$$

Let

$$\widetilde{\pi} = (a_1 \pi^1, a_2 \pi^2, \ldots, a_e \pi^e),$$

then

$$\widetilde{\pi}_i > 0, \qquad \text{for} \quad i = 1, \ldots, K$$

and

$$\begin{aligned}\sum_{i=1}^{K} \tilde{\pi}_i &= \sum_{i=1}^{e}\sum_{j=1}^{e_i} a_i \pi_j^i \\ &= \sum_{i=1}^{e} a_i \\ &= 1.\end{aligned}$$

Moreover,

$$\begin{aligned}\tilde{\pi}A &= (a_1\pi^1 B_1, a_2\pi^2 B_2, \dots, a_e\pi^e B_e) \\ &= (a_1\pi^1, a_2\pi^2, \dots, a_e\pi^e) \\ &= \tilde{\pi},\end{aligned}$$

then $\tilde{\pi} \in S^+$ and hence $S^+ \neq \emptyset$.

To prove (c), let $\{X_t\}$ be a Markov chain having k communicating classes, $2 \leq k \leq K$, where e of them are ergodic, $1 \leq e < k$, and t of them are transient, $e + t = k$. Let K_e and K_t be the number of ergodic states and transient states of $\{X_t\}$ respectively. Without loss of generality, assume that the matrix transition A is of the form

$$A = \begin{pmatrix} B & 0 \\ C & D \end{pmatrix},$$

where B is a $K_e \times K_e$ stochastic matrix, D is a $K_t \times K_t$ substochastic matrix and C is a $K_t \times K_e$ matrix, $C \neq 0$.

Let $\pi = (\pi_1, \dots, \pi_K) \in S$, since

$$\pi A = \pi,$$

then

$$\begin{aligned}A^T\pi^T &= \pi^T \\ (A^T - I_K)\pi^T &= 0\end{aligned}$$

or

$$\begin{pmatrix} B^T - I_{K_e} & C^T \\ 0 & D^T - I_{K_t} \end{pmatrix}\begin{pmatrix} \pi^1 \\ \pi^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{13}$$

where $\pi^1 = (\pi_1, \dots, \pi_{K_e})$ and $\pi^2 = (\pi_{K_e+1}, \dots, \pi_K)$. By (13), π^1 and π^2 satisfy

$$(B^T - I_{K_e})\pi^1 + C^T\pi^2 = 0 \tag{14}$$

$$(D^T - I_{K_t})\pi^2 = 0. \tag{15}$$

By [2], page 44, $D^T - I_{K_t}$ is invertible, so the only solution for (15) is $\pi^2 = 0$. So π must have form $\pi = (\pi^1, 0)$, where π^1 satisfy (14). This means that $\pi \in S^o$. Therefore $S \subset S^o$, implying $S = S^o$ and $S^+ = \emptyset$. ∎

References

[1] F.R. Gantmacher, *Applications of the Theory of Matrices.* New York: Interscience, 1959.

[2] J.G. Kemeny and J.L. Snell. *Finite Markov Chains*, New York: Springer Verlag, 1976.